Figures and figure supplements

Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections

Christopher JR Illingworth *et al*

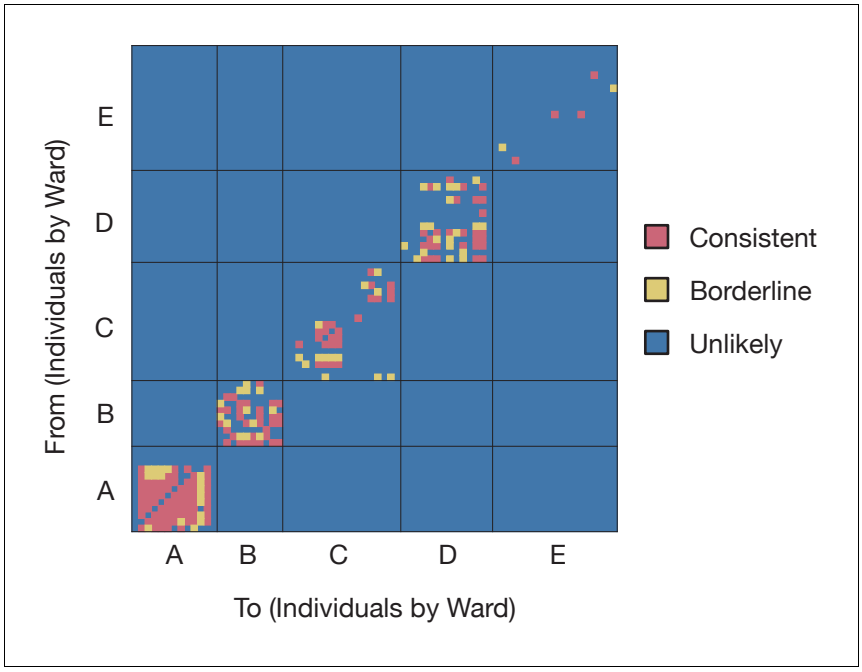

Figure 1. Preliminary analysis of the data with A2B-Covid. Squares indicate the extent to which an individual-to-individual transmission event is consistent with the data collected, when considered on a pairwise level. Our analysis highlighted multiple potential transmission events occurring within each ward, but transmission between individuals on different wards was uniformly assessed as unlikely. Further analyses of the data considered wards as independent and isolated locations.

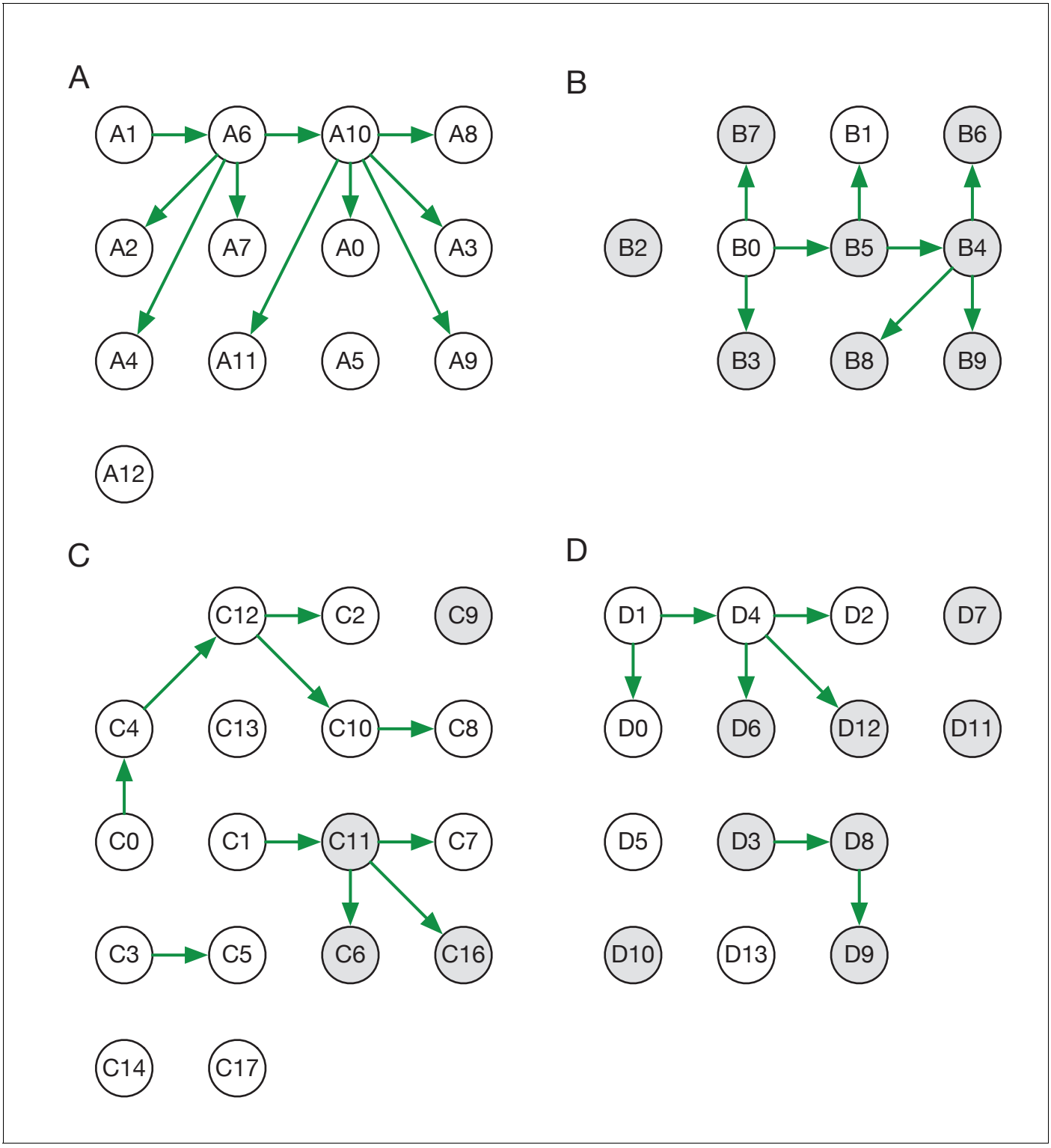

Figure 2. Maximum likelihood transmission networks for wards A to D. Circles represent individuals and arrows show transmission events. White circles represent patients while grey circles represent health care workers. Individuals for which no transmission events were inferred are represented as isolated circles.

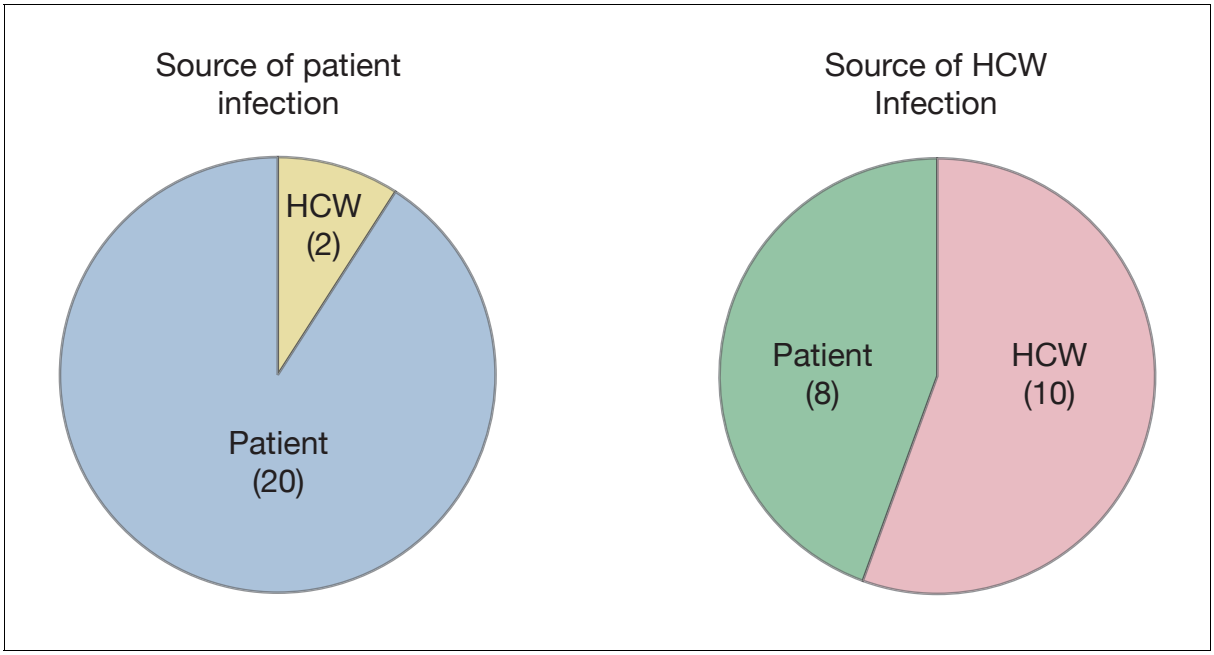

Figure 2—figure supplement 1. Maximum likelihood sources of patient and HCW infections. Statistics were calculated across maximum likelihood network reconstructions. The great majority of patient infections were inferred to arise from other patients, while HCWs were infected roughly equally by patients and other HCWs.

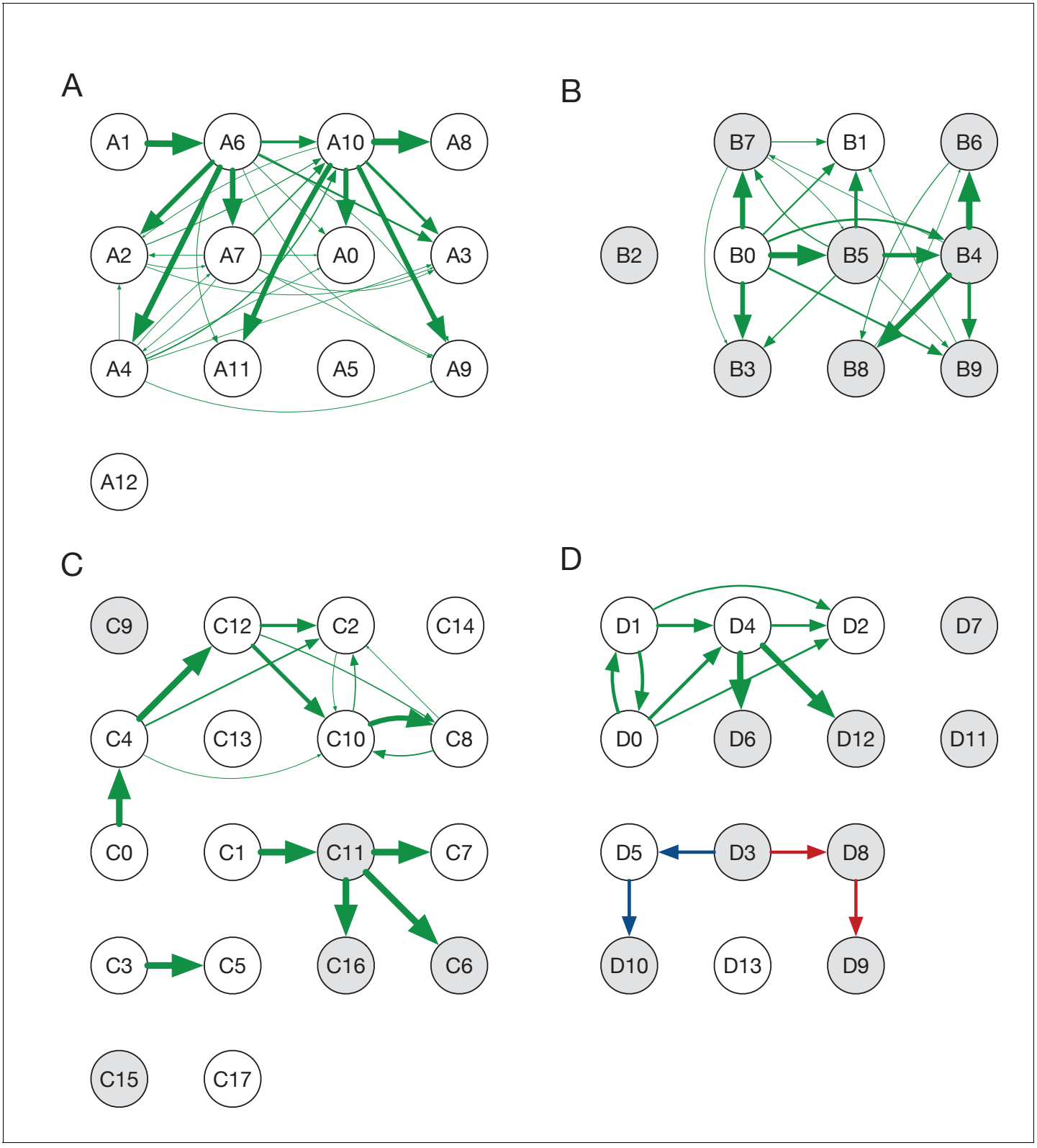

Figure 2—figure supplement 2. Ensemble transmission networks for wards A to D. Data were compiled over sets of plausible reconstructions, weighted by likelihood. The width of each arrow is proportional to the probability that a specific transmission event occurred. Arrows are shown for events with a 4% or greater probability of having occurred. Red and blue lines indicate mutually incompatible events; D3 could have infected D5 or D8, but the data precluded both of these occurring in the same reconstruction.

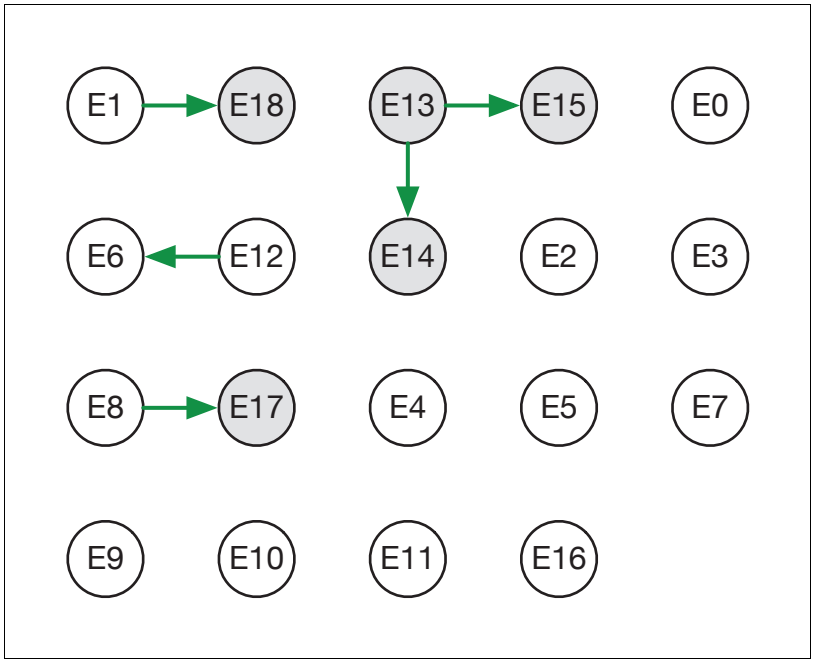

Figure 3. Maximum likelihood transmission network for ward E. Circles represent individuals and arrows show transmission events. White circles represent patients while grey circles represent health care workers. Individuals for which no transmission events were inferred are represented as isolated circles.

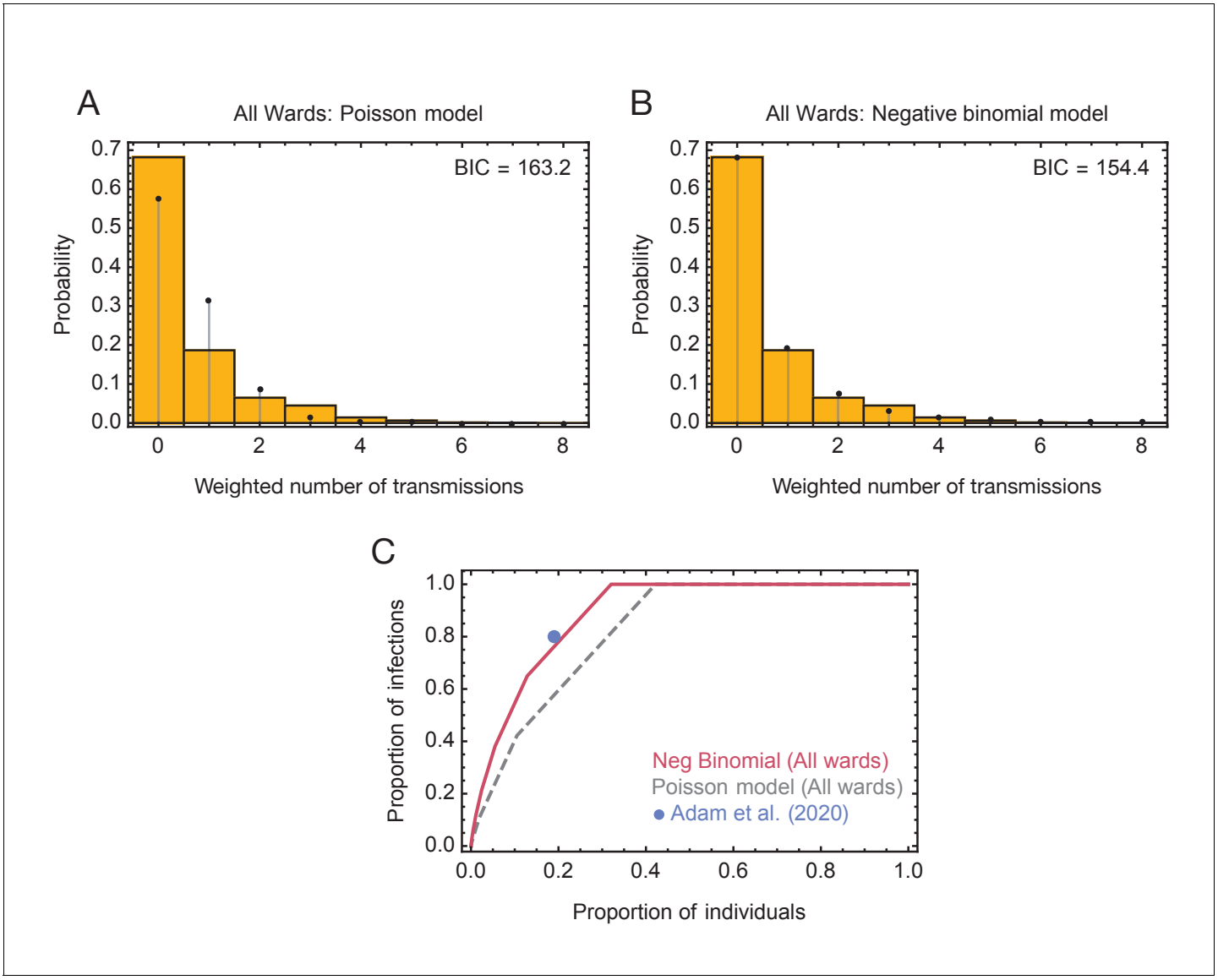

Figure 4. Models of viral transmission. (**A**) Fit of the output of the Poisson model (black dots) to the ensemble data (yellow bars). The weighted number of transmissions per individual reflects the uncertainty in the network reconstruction across the ensemble. (**B**) Fit of the output of the negative binomial model (black dots) to the ensemble data (yellow bars). (**C**) Proportions of individuals causing different proportions of infections. A negative binomial model (red line) fitted to all ward data produces a result similar to that of ***Adam et al., 2020*** (blue dot), with 20% of individuals being responsible for 80% of infections. A Poisson model fitted to the same data (dashed grey line) has 20% of individuals being responsible for 60% of infections.

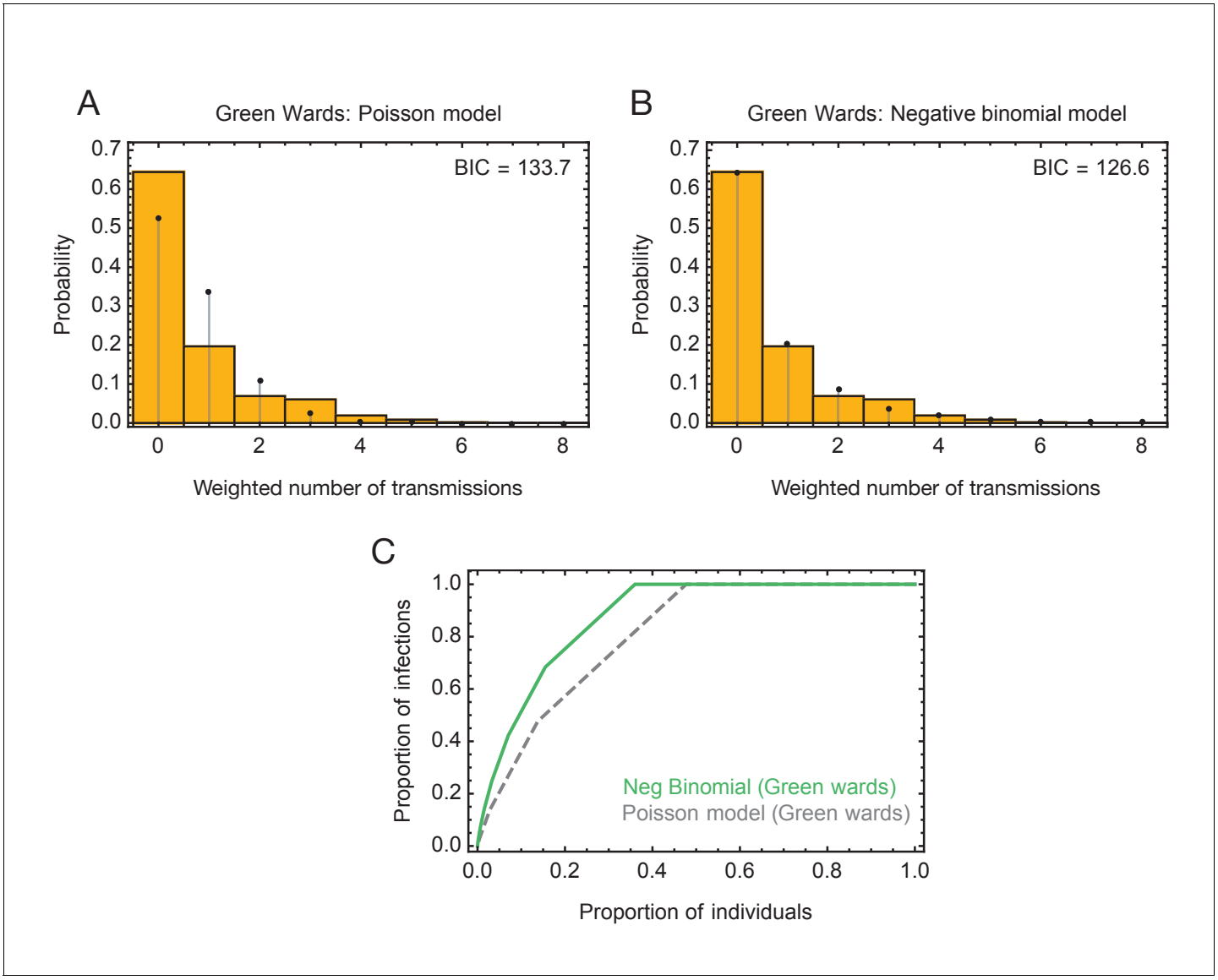

Figure 4—figure supplement 1. Modelling of viral transmission on the green wards. (**A**) Fit of the output of a Poisson model (black dots) to the ensemble data (yellow bars). The weighted number of transmissions per individual reflects the uncertainty in the network reconstruction across the ensemble. (**B**) Fit of the output of the negative binomial model (black dots) to the ensemble data (yellow bars). (**C**) Proportions of individuals causing different proportions of infections. A negative binomial model (green line) fitted to data from the green wards suggests that 20% of individuals were responsible for 75% of infections. A Poisson model fitted to the same data (dashed gray line) has 20% of individuals being responsible for 58% of infections.

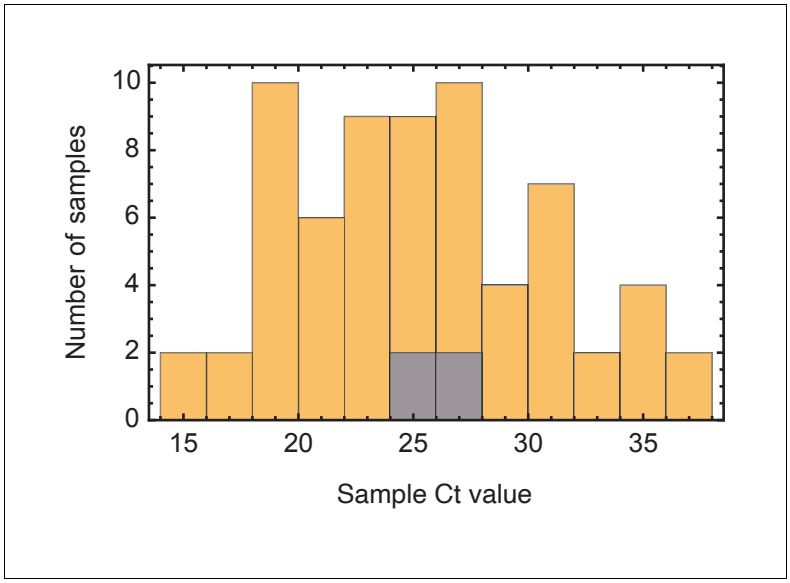

Figure 4—figure supplement 2. Ct values of viral samples. Distributions of known Ct values collected from all samples from the wards studied (yellow) and from individuals identified as superspreaders (grey). Samples from superspreaders were not statistically different from those from the population as a whole. Ct values were not available for all samples.

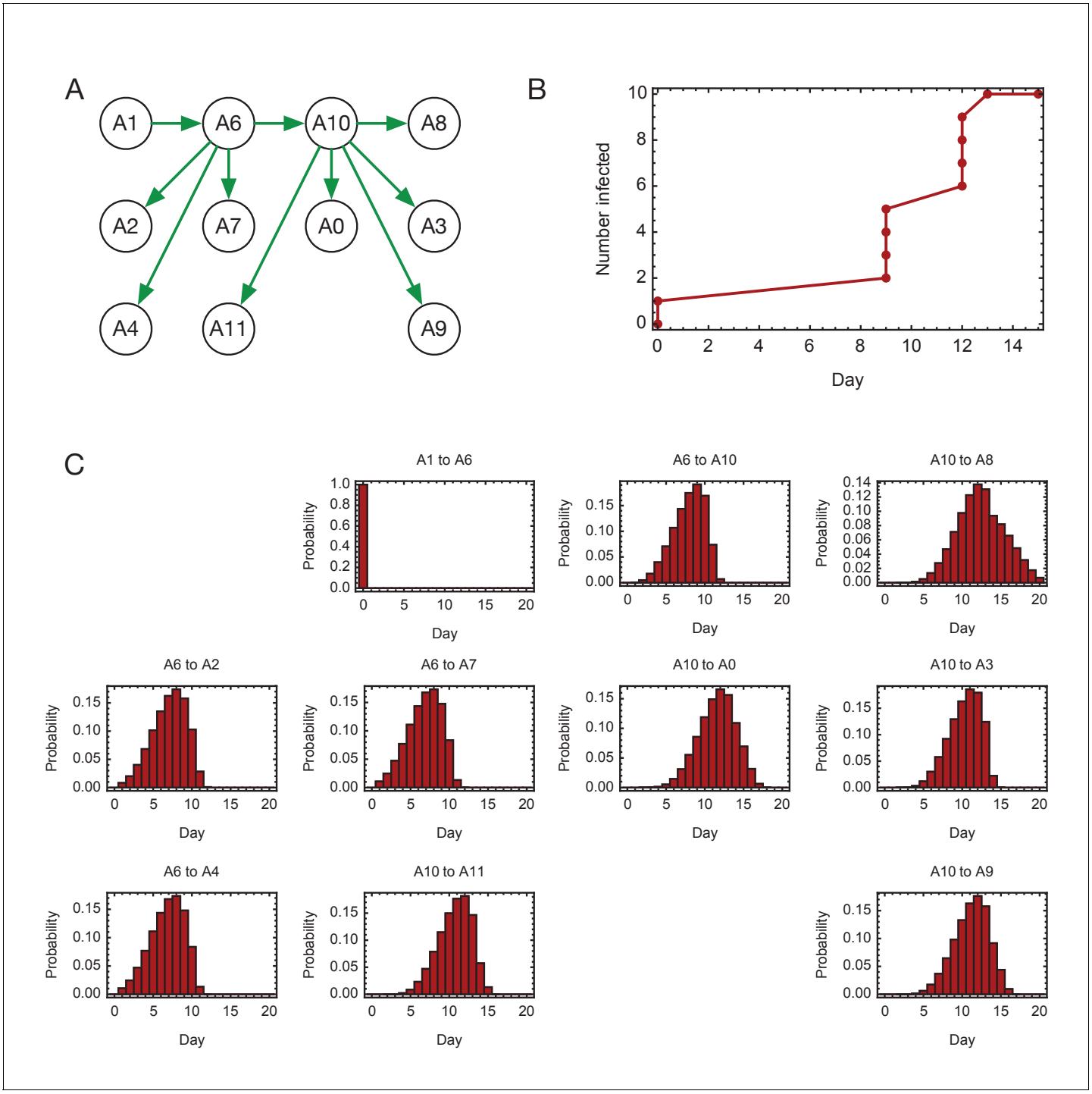

Figure 4—figure supplement 3. Inferred timings of transmission events in ward A. (**A**) Maximum likelihood network of transmission events. (**B**) Maximum likelihood spread of infection given this network. (**C**) Distributions of the times at when transmission occurred were calculated relative to the time of the first transmission event, in this case from A1 to A6. Timings account for infection dynamics and the locations of individuals during the course of the outbreak.

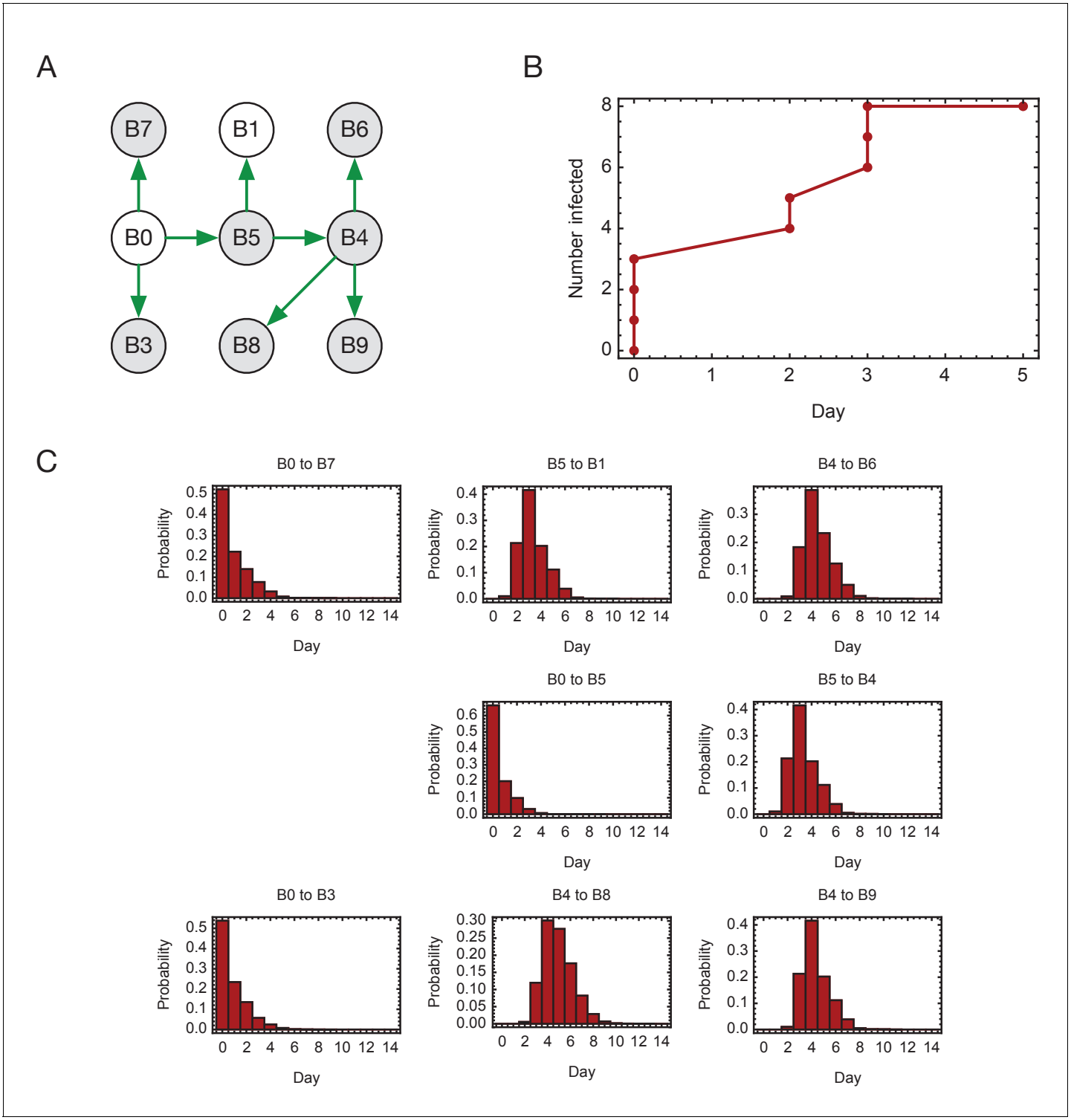

Figure 4—figure supplement 4. Inferred timings of transmission events in ward A. (**A**) Maximum likelihood network of transmission events. (**B**) Maximum likelihood spread of infection given this network. Timings are shown relative to the first transmission event. Infection of the first individual (B0) is not included. (**C**) Distributions of the times at when transmission occurred were calculated relative to the time of the first transmission event. Timings account for infection dynamics and the locations of individuals during the course of the outbreak.

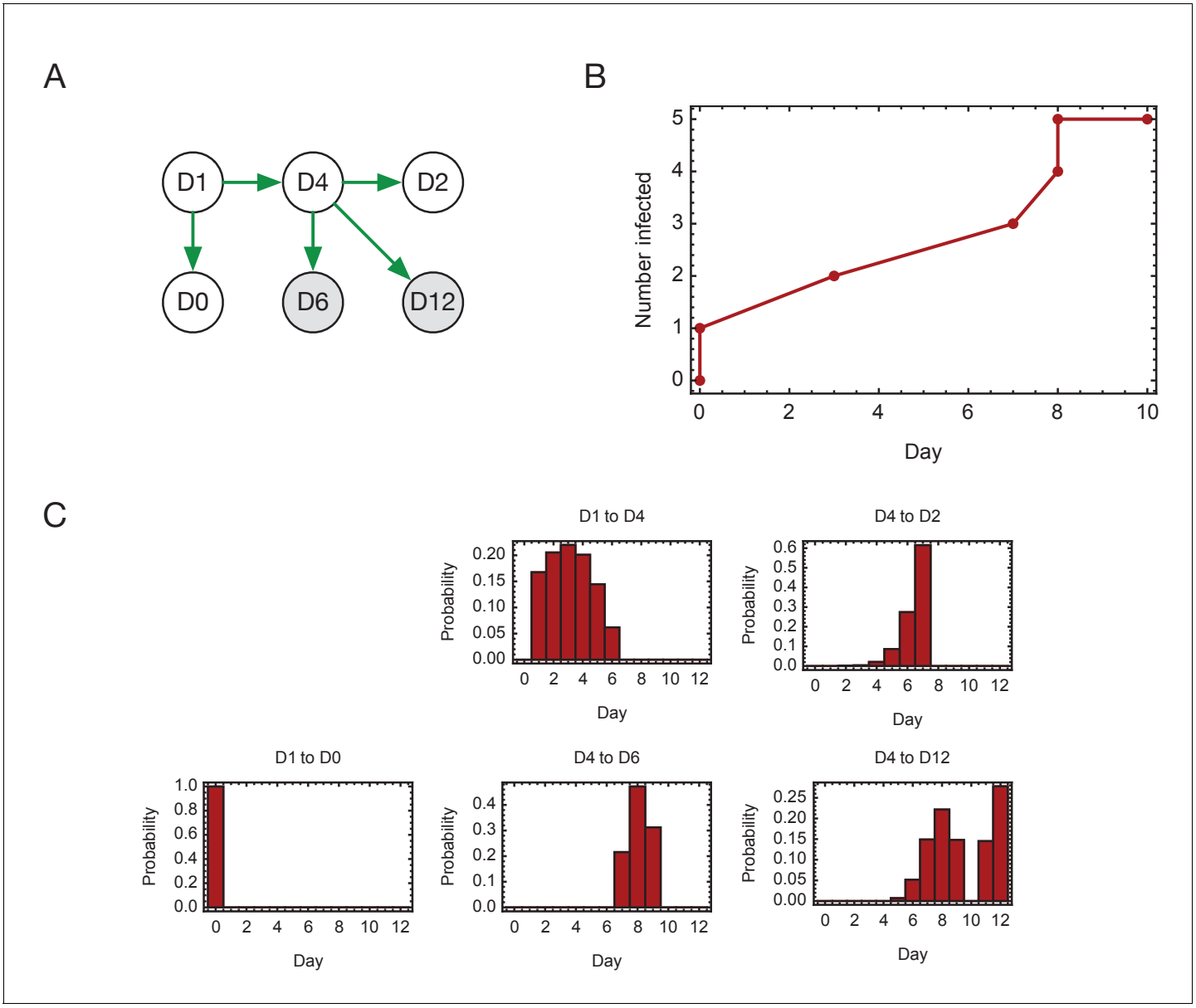

Figure 4—figure supplement 5. Inferred timings of transmission events in ward D. (**A**) Maximum likelihood network of transmission events. (**B**) Maximum likelihood spread of infection given this network. Timings are shown relative to the first transmission event. Infection of the first individual (D1) is not included. (**C**) Distributions of the times at when transmission occurred were calculated relative to the time of the first transmission event. Timings account for infection dynamics and the locations of individuals during the course of the outbreak.

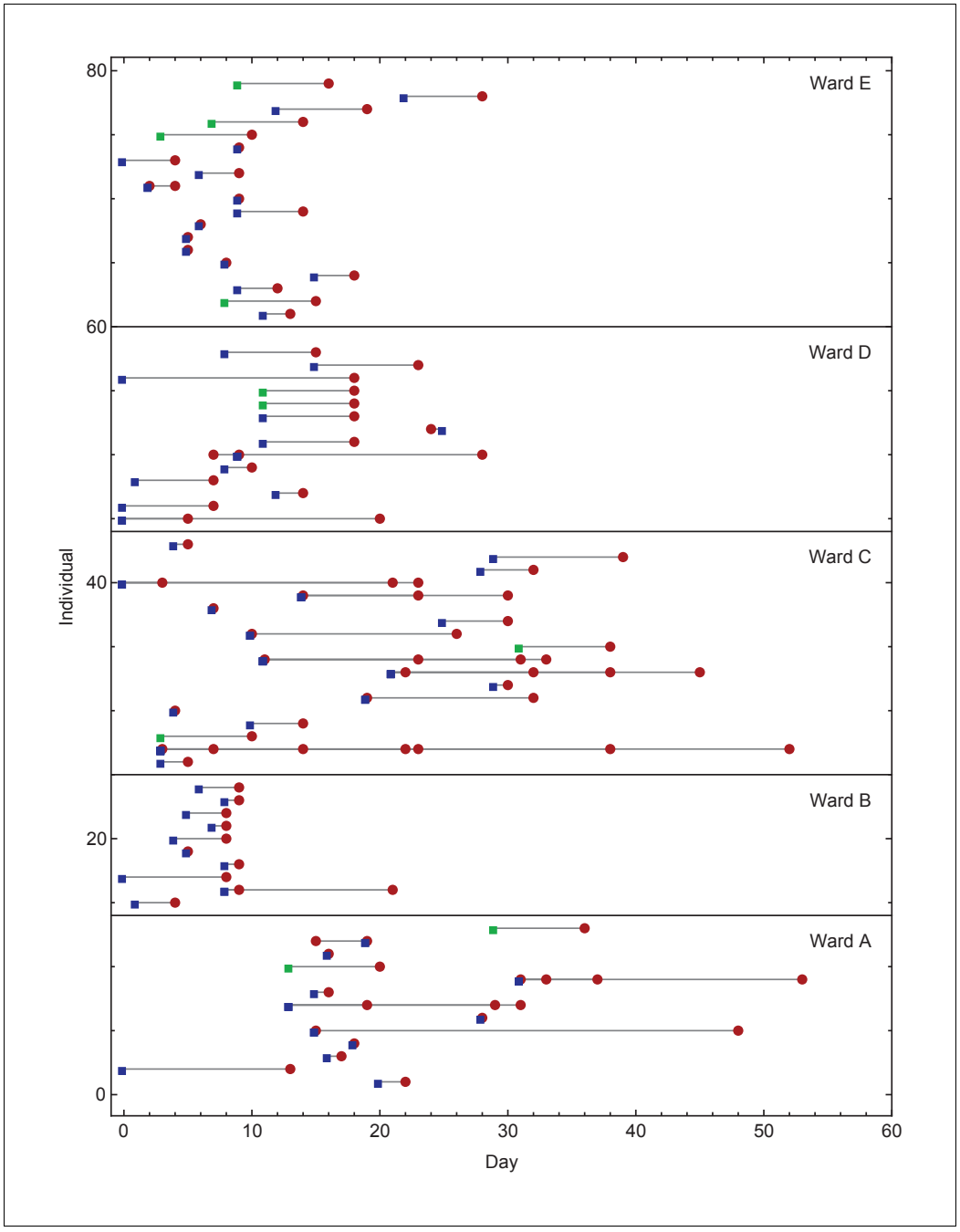

Figure 5. Overview of events on different wards. Blue squares show days on which individuals became symptomatic, while green squares show inferred days of individuals becoming symptomatic when these dates were unknown or not applicable. Red circles show days on which samples were collected from individuals for genome sequencing. Dates within each ward are normalised so that the first event on any ward is day zero. We note that not all collected samples led to genome sequences of sufficient quality to be useful in this study.

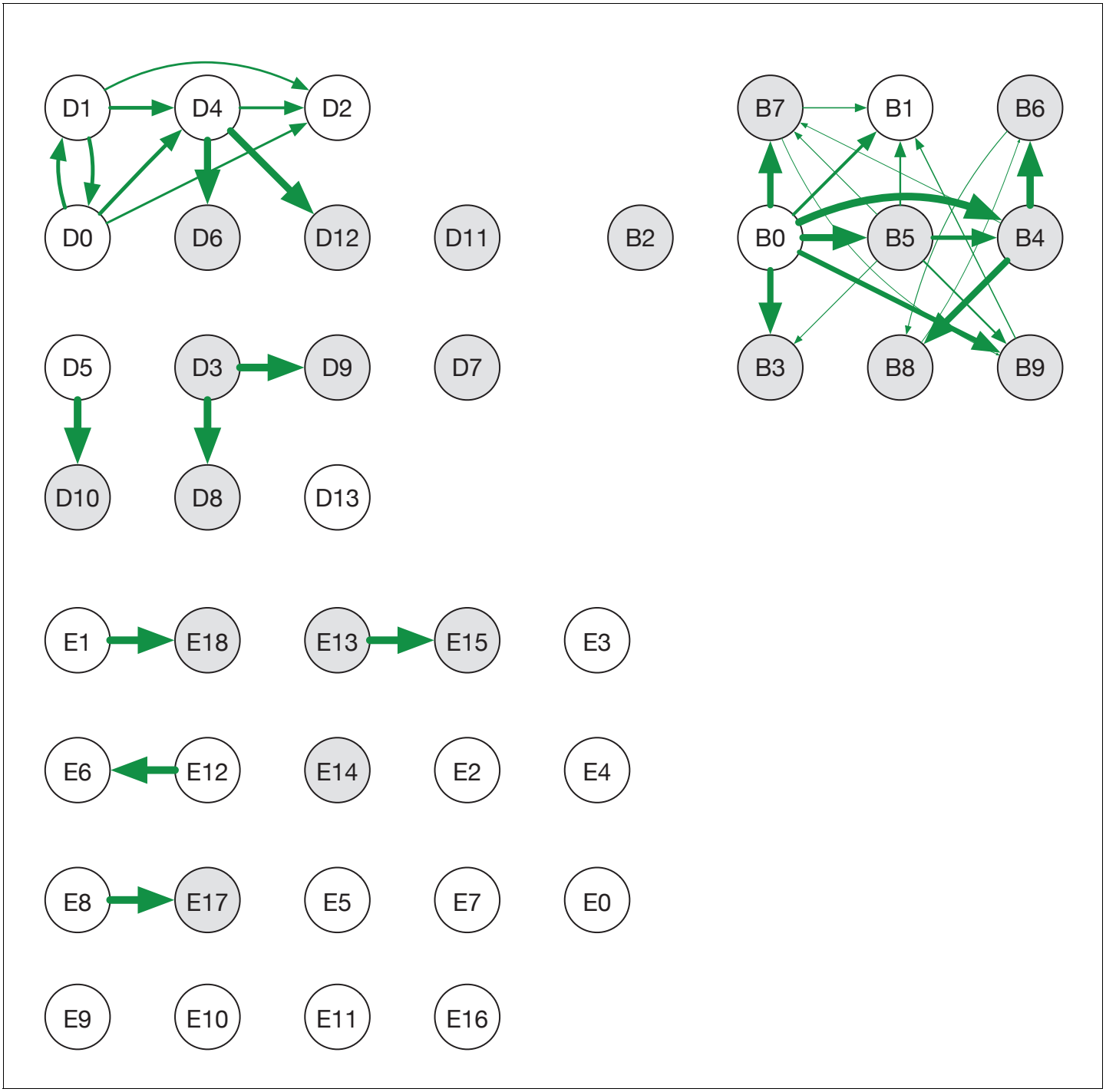

Figure 5—figure supplement 1. Ensemble transmission networks for wards B, D, and E generated without extending the times at which HCWs were present beyond the direct observations made. Data were compiled over sets of plausible reconstructions, weighted by likelihood. The width of each arrow is proportional to the probability that a specific transmission event occurred. Arrows are shown for events with a 4% or greater probability of having occurred. Ensembles for wards A and C were not sufficiently changed by removing the extension for the change to be visible within our plotting framework.

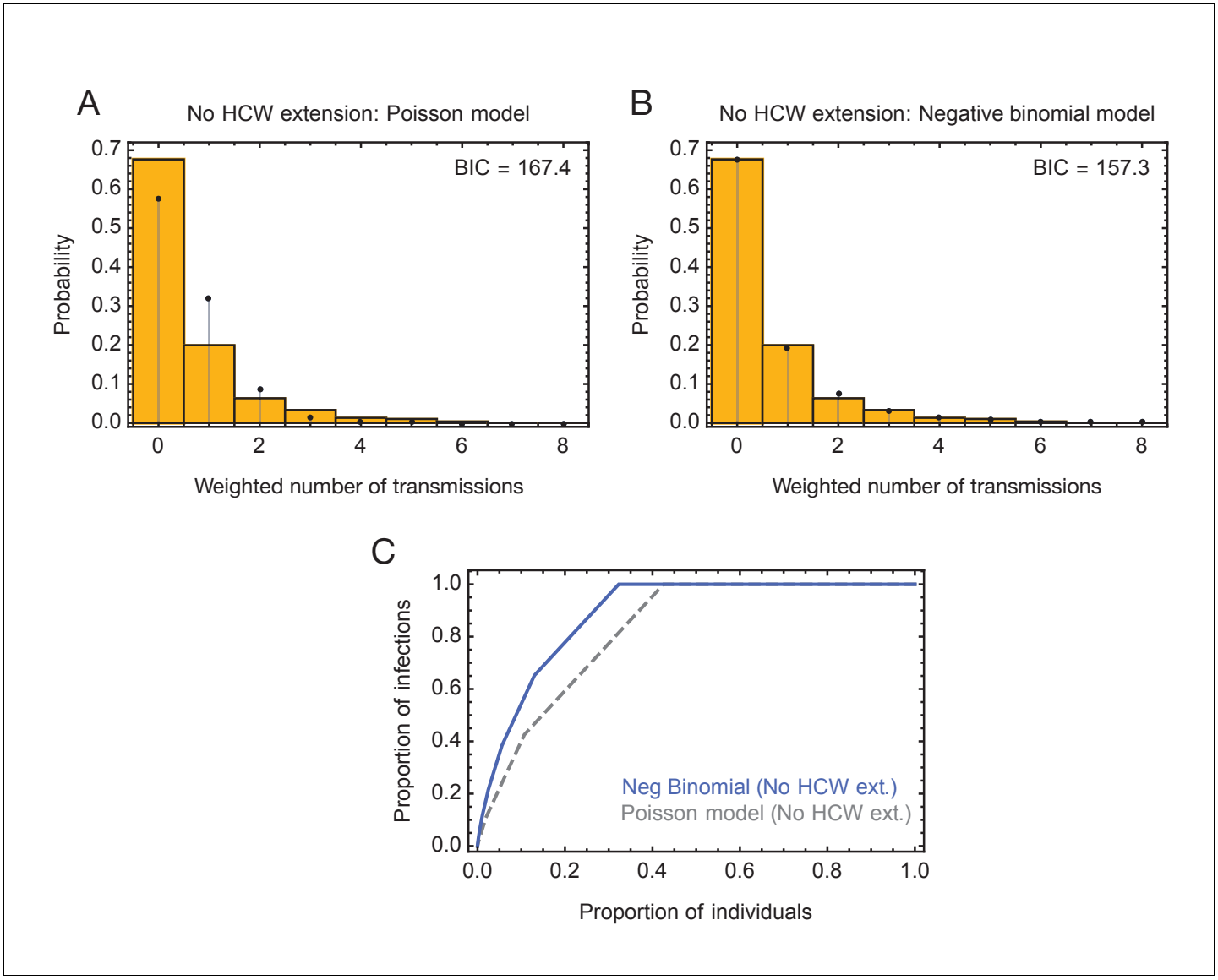

Figure 5—figure supplement 2. Modelling of viral transmission in the absence of an extension to HCW locations. (**A**) Fit of the output of a Poisson model (black dots) to the ensemble data (yellow bars). (**B**) Fit of the output of the negative binomial model (black dots) to the ensemble data (yellow bars). (**C**) Proportions of individuals causing different proportions of infections. A negative binomial model (blue line) fitted to data from the green wards suggests that 21% of individuals were responsible for 80% of infections.

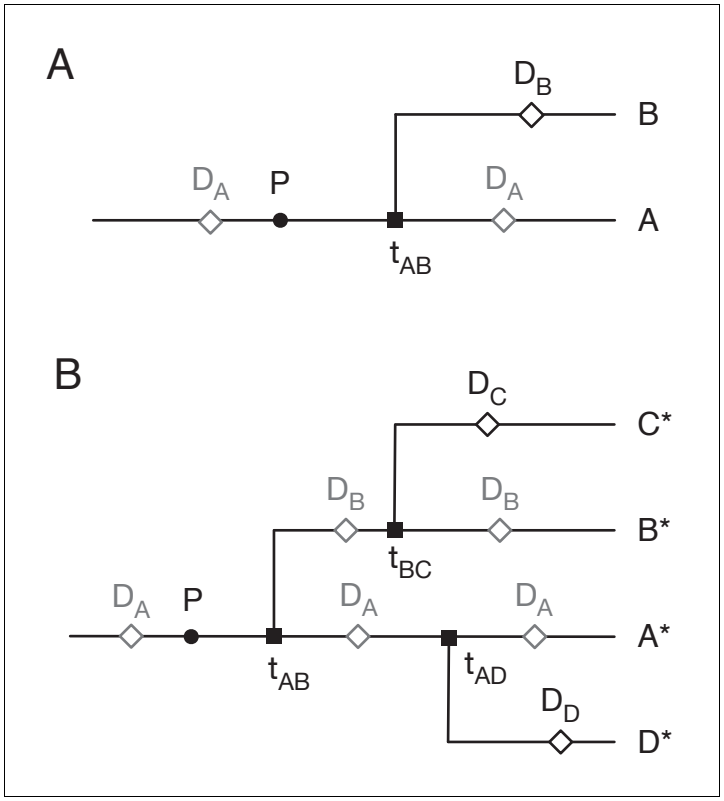

Figure 5—figure supplement 3. Assigning mutations to the transmission tree. (**A**) Case of the last transmission event. Transmission occurs from A to B at time t_{AB}. Viral sequence data is collected from A at time D_A and from B at time D_B. Grey markers show points which may be located in multiple places on the tree. (**B**) General case of transmission from A to B. The notation A* denotes a lineage that includes the individual A plus potentially further individuals downstream to whom A transmits the virus.

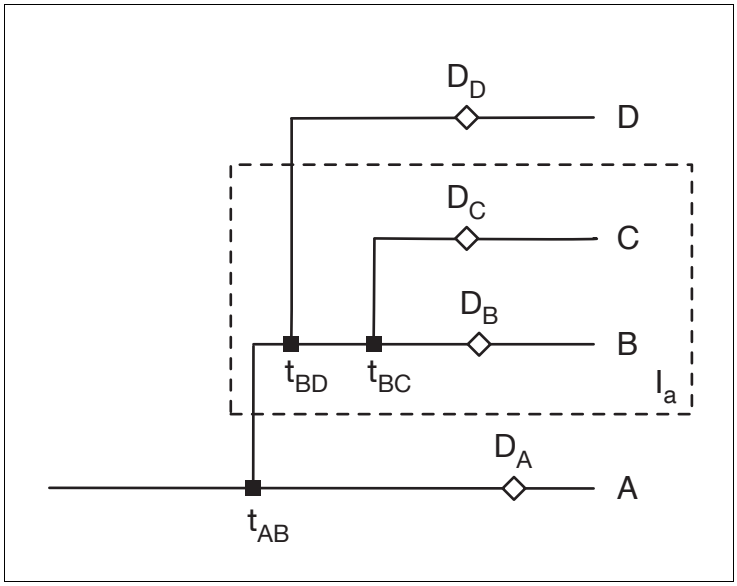

Figure 5—figure supplement 4. Restrictions placed on the network by sequence variants. Here sequences collected from the individuals B and C (i.e. in the set I_a) have the variant a, but no other sequences have this variant. Data from individual i was collected at time D_i. We assume that variants can be gained only once, and that variants never revert. Then 1: There can be only one transmission into the set I_a. Suppose that A, who is not in Ia, transmits to B, who is in I_a. Then the gain of the variant must occur between the earlier of t_{AB} and D_A and the latter of D_B and t_{BC}. 2: B can transmit to D not in I_a, but no other individual in I_a can transmit out of I_a. Transmission from B can occur before the gain of the variant, but transmission from any other C in I_a would involve the reversion of the variant. 3. All transmissions from B to D not in I_a must occur before all transmissions from B to C in I_a.

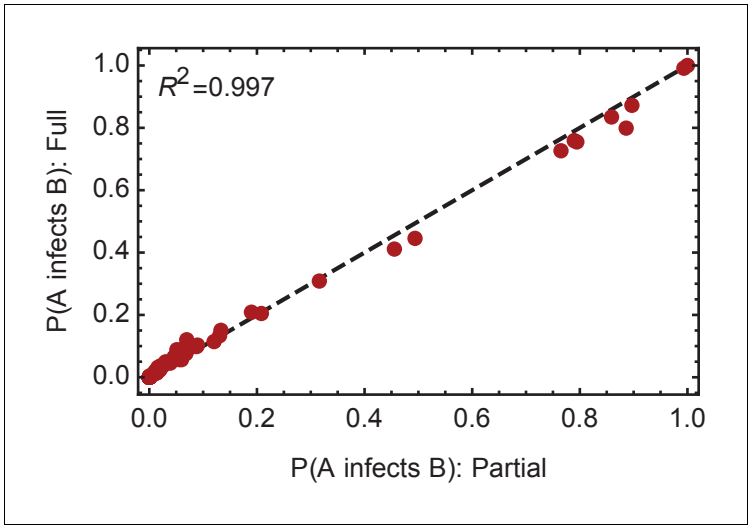

Figure 5—figure supplement 5. Convergence of the statistical ensemble of networks for ward A. Comparisons of the number of infections per individual and the probabilities of specific edges being found in the transmission network for a 'partial' set of networks and for a more complete, 'full' set of networks. The full set contains approximately 30% more networks than the partial set, adding in likelihoods calculated for networks within three steps of the maximum likelihood network. Statistics calculated over the two sets are extremely similar, suggesting convergence to the true statistical ensemble.